Why Do Defendants Ever Plead Guilty?

Otto H. Swank

(Erasmus School of Economics / Tinbergen Institute)

September 3, 2011

Abstract

At the beginning of a criminal trial, the defendant is asked to plead guilty or not guilty. This paper shows that the pleading decision serves as a signaling device. Guilty defendants to whom mitigating circumstances apply plead guilty, while innocent defendants and defendants to whom aggravating circumstances apply plead not guilty. With a view of imposing proper sentences, too many defendants may plead guilty. The reason is that the plea decision contains information about the proper sentence.

JEL Classification: D73, D82, K1
Keywords: guilty pleas, information collection

Corresponding address:
Otto H. Swank
Erasmus Univeristy Rotterdam,
P.O. Box 1738
3000 DR, Rotterdam
The Netherlands
E-mail: swank@ese.eur.nl

1 Introduction

In most countries, criminal trial courts operate under an adversarial system, in which two advocates, the prosecutor and the defendant's attorney, represent their parties positions before a neutral person or group of persons. Theoretical models show both an advantage and a drawback of the adversarial system (see e.g. Sobel, 1985, Milgrom and Roberts, 1987, Dewatripont an Tirole, 1999, and Kim, 2010a). The advantage is that the adversarial system provides strong incentives to each advocate to search for evidence favoring his cause. A drawback of the adversarial system is that it also gives incentives to each advocate to conceal information that may hurt his cause.

The existing theoretical literature on adversarial systems sheds light on the incentives lawyers have. However, one prediction of this literature seems to be inconsistent with actual practices. At the beginning of a criminal trial process, the defendant is asked to plead guilty or not guilty. The inconsistency is that pleading guilty is quite common (also in countries where little room for plea bargaining exists). This brings me to the main question of the present paper: why do defendants sometimes plead guilty? I believe that this question is the more intriguing as in criminal trials the prosecutor has the burden of proof. That is, if the prosecutor is not able to present evidence that the defendant is guilty, the defendant will not be convicted.

To answer my main question, I develop a game-theoretical model of a criminal trial. The game describes the behavior of the defendant's attorney, the prosecutor, and the court. The model has five main (related) features. First, at the beginning of the game, the defendant either pleads guilty or not guilty. Second, the court has two tasks. It has to determine whether the defendant is guilty or not, and, if the defendant has been found guilty, it has to impose a sentence. Third, I distinguish between two kinds of information. One kind of information concerns evidence about the crime in relationship with the defendant. The other kind of information is related to the gravity of the crime.[1] This information is important for the proper sentence if the defendant has been found guilty. The idea is that mitigating or aggravating circumstances may apply. Fourth, I assume that information is asymmetric. The

[1] For example, one can think of the defendant's mental state (mensua) or attendant circumstances.

defendant's attorney knows whether his client is guilty and, if guilty, which sentence he should get. The prosecutor, by contrast, only knows prior distributions. With some exogenous probability, the defendant's attorney is able to make his information verifiable. The prosecutor learns information about the crime and circumstances with exogenous probabilities. Finally, the prosecutor has the burden of proof. If she does not put forward evidence that the defendant is guilty, the defendant will not be convicted.

My main result is that the pleading decision serves as a signaling device. Two types of defendants plead not guilty, the innocents and the defendants to whom aggravating circumstances apply. Guilty defendants to whom mitigating circumstances apply plead guilty. The intuition behind this result is straightforward. A guilty defendant who risks a harsh sentence has no incentive to plead guilty. Pleading guilty leads to a certain conviction and thereby increases the probability of a harsh sentence. A guilty defendant to whom mitigating circumstances apply faces a trade-off. Pleading guilty leads to a certain conviction on the one hand, but reduces the expected sentence on the other. As defendants with mitigating circumstances plead guilty, pleading guilty signals that one deserves a relatively mild sentence.

Once I have shown that defendants may plead guilty, I examine the welfare consequences of allowing defendants to plead guilty. Obviously, guilty pleas increase conviction rates. In general, this is good from a social perspective. However, as the plea decisions serve as a signaling device, there is a cost of too high a proportion of guilty pleas. If all guilty defendants plead guilty, the plea decision does not longer contain information about the gravity of the crime. In my model, from the optimal sentence point of view, half of the guilty defendants should plead guilty.

Finally, I discuss two potential drawbacks of allowing defendants to plead guilty. First, innocent persons may plead guilty. Of course, from a social perspective, it is undesirable that innocent defendants are convicted. However, my model suggests that an innocent person only pleads guilty if he perceives a sufficiently high probability that he will be convicted in case he pleads not guilty. Given this high probability, the possibility to plead guilty makes the (innocent) defendant better off. Second, the possibility to plead guilty may reduce deterrence.

This paper is related to two strands in the literature. The literature on plea bargaining is most closely related.[2] In the United States, most convictions in criminal cases are the result of plea-bargaining where the prosecutor offers the defendant a relatively low sentence in exchange for a guilty plea. Grossman and Katz (1983) show that plea bargaining can serve as a screening device (see also Reinganum, 1988). The idea is that guilty defendants prefer a conviction with a relatively mild sentence to a risk of a conviction with a relatively harsh sentence. In Grossman and Katz only innocent people proceed to trials. Baker and Mezzetti (2001) point to a weakness in the Grossman and Katz model. If only the innocent defendants proceed to trial, why should a prosecutor go to a time-consuming trial when he knows that the defendant is innocent? Clearly, without commitment, the prosecutor would drop the case. However, this would give guilty defendants an incentive to reject the prosecutor's offer and to go for a trial. Baker and Mezzetti show that when you relax the assumption of commitment, some guilty defendants accept the prosecutor's plea proposal, while others do not. As a guilty defendant compares the expected benefit of accepting a proposal to the expected benefit of not expecting the proposal, in Baker and Mezzetti, the defendants risking heavier sentences particular benefit from accepting a plea proposal.

Our model deviates in two main respects from the game-theoretical literature on plea bargaining. First, a plea bargaining is similar in form as a voluntary transaction between two parties (Grossman and Katz, 1983). The plea decision is not a transaction. It is just a statement of one player. Second, as discussed above, in the plea bargaining literature, the guilty defendants to whom aggravating circumstances apply have strongest incentives to enter in plea bargaining. For them, the risk of a trial is highest. In my model, the defendants to whom mitigating circumstances apply plead guilty.

My paper is also related to the literature on information collection and communication. Milgrom and Roberts (1986) show that when an interested party knows the true state of the world and information is verifiable, he always has an incentive to reveal his information to the decision maker. The reason for this result is that if a decision maker were not to receive information, she would assume the worst. Likewise, when two interested parties with opposing objectives possess information, one

[2] See for a very nice survey of the literature on the political economy of prosecution Gordon and Huber (2009).

of them has an incentive to reveal information to the decision maker. Full revelation disappears when parties are not always informed (see, for example, Austen Smith, 1994, and Shin, 1994). In that case, a party that possesses information may have an incentive to pretend to have no information. This pooling of uninformed and informed agents may also take place in my model. Dewatripont and Tirole (1999) show that parties with opposing preferences, advocates, have strong incentives to collect information (see also Dur and Swank, 2005). Dewatripont an Tirole provide a rationale for an adversarial system (see also Kim, 2010b, and Posner, 1999). My paper builds on the above literature on information collection and transmission. As far as I know, the present paper is the first that can explain why in an adversarial system, an interested player may provide evidence against himself.

2 A Simple Model

I consider a situation where a person is accused of having committed an illegal act. I refer to this person as the suspect or defendant. Two characteristics of the suspect in relationship with the act are important. First, whether the suspect actually committed the act, $x = 1$, or did not commit the act, $x = 0$. The prior probability that the suspect is guilty is ρ, $\Pr(x = 1) = \rho$. Second, the external circumstances. These are captured by a stochastic term μ, which is uniformly distributed on $[l, h]$, with $l \geq 0$. The external circumstances are only relevant if the suspect committed the act. Circumstances can be mitigating. This is associated with low values of μ. They can also be aggravating (high values of μ). From a social point of view it is optimal that a guilty person gets a sentence $s = \mu$.[3]

The suspect is brought into court. The court has to make two decisions. First, it has to determine whether the suspect is guilty or not, $g \in \{0, 1\}$ with $g = 1$ denoting guilty and $g = 0$ denoting not guilty. I assume that the court chooses $g = 1$ if it learnt that $x = 1$. Second, if $g = 1$, the court must choose a sentence, s. I assume that if $g = 1$ the court chooses s according to $s = E(\mu | I_C)$ where I_C is the information the court possesses when choosing s. Obviously, if $g = 0$, $s = 0$. My

[3] External circumstances can refer to the gravity of the crime or to personal characteristics of the defendant. What matters for the present paper is that from a social point of view the sentence depends on μ. As a result, imposing the proper sentence is a real task of the court.

assumptions imply that the court needs hard evidence for a conviction. However, concerning the sentence the court is assumed to be a Bayesian updater.

Apart from the court, there are two agents $i \in \{D, P\}$. First, the defendant's attorney (D, he) wants the suspect, his client, not to be convicted. If his client is proved guilty, he tries to persuade the court to impose the lowest sentence possible. D's preferences are described by the following payoff function

$$U_D\left(g, s\right) = -\lambda_D g - s$$

where $\lambda_D > 0$ denotes the cost D attaches to $g = 1$. Second, the prosecutor (P, she) tries to convince the court that the suspect is guilty. I assume that P wants the court to impose the highest sentence possible. P's payoff function is

$$U_P\left(g, s\right) = \lambda_P g + s$$

where $\lambda_P > 0$ denotes the benefit P receives from $g = 1$. Note that D and P have opposite interests.

The model tries to capture some main features of a criminal trial. I assume that D knows x and μ, but that the court and P do not know x and μ. At the beginning of the game, D can either admit that his client is guilty or not, $x_D \in \{0, 1\}$, where $x_D = 0$ denotes plead not guilty, and $x_D = 1$ denotes plead guilty. x_D is a cheap talk message to both P and the court. $x_D = 1$ always induces the court to choose $g = 1$.[4] We assume that P has the burden of proof. This means that if $x_D = 0$, for a conviction P must present evidence that the defendant is guilty.

After P and the court have received message x_D, D and P collect information. As described above, D knows x and μ. In my model if $g = 1$, D wants the court to believe that μ is small. I assume that with probability $\pi_D^\mu \in (0, 1)$, D is able to present hard evidence about μ. If D were to convey such evidence to the court and $g = 1$, then the court would choose $s = \mu$. If $x = 1$, with probability $\pi_P^x \in (0, 1)$, P finds hard evidence that the suspect is guilty, and with probability $1 - \pi_P^x$ she does not find evidence that the suspect is guilty. If $x = 0$, P cannot find evidence that the suspect is guilty (in Section 5 I relax this assumption). With probability

[4] Until Section 5, I abstract from the possibility of false confessions.

$\pi_P^\mu \in (0,1)$, P finds hard evidence on μ. Note that the probabilities with which P finds information is independent of D's pleading decision. Below I show that this does not mean that the pleading decision is independent of the probabilities with which P finds hard evidence. It does mean that the pleading decision is not driven by a desire to affect the probability that P learns μ.

Next, after the information collection stage, D and P may convey information to the court. I call this the communication stage of the game. If $x_D = 0$ and P cannot present evidence that the suspect is guilty, then the court chooses $g = 0$ and $s = 0$, and the game ends. If $x_D = 1$, or $x_D = 0$ and P has presented evidence that the suspect is guilty, both D and P may present evidence about μ. I model this by assuming that D and P simultaneously send a message about μ, where $m_i = \varnothing$ denotes that i presents no evidence, and $m_i = \mu$ denotes that i reveals μ. So, either D or P may reveal information or conceal it.

At the end of the game, the court makes a decision on x and s. The structure of the game, the information structure, and the prior beliefs are all common knowledge.

3 Equilibria

My game is a dynamic game with incomplete information. I solve it by backward induction. I identify Perfect Bayesian equilibria, in which, given beliefs, strategies are optimal responses to each other, and beliefs are updated according to Bayes' rule, whenever possible.

I split the analysis into two parts. First, I assume that concerning the pleading decision, D follows a threshold strategy: plead guilty ($x_D = 1$) if $x = 1$ and $\mu < \mu_x$, and plead not guilty otherwise. Assuming this strategy, I solve the communication games. Next, I focus on the pleading decision, assuming that D anticipates equilibrium behavior in the communication stage.

3.1 The Persuasion Game

Assume that $x_D = 1$ if and only if $x = 1$ and $\mu < \mu_x$. Recall that if $x_D = 0$ and P does not find evidence that $x = 1$, the game ends. For the communication stage, two cases remain. First, $x_D = 1$. Then, all players know that $l \leq \mu < \mu_x$. Second, $x_D = 0$ and P has presented evidence that $x = 1$. Then, all players know that

$\mu_x \leq \mu < h$. Note that in both cases the court knows that the defendant is guilty. The trial revolves around the sentence.

Consider the first case. I will identify an equilibrium in which both D and P follow a threshold strategy: if informed, D reveals information if and only if $\mu < \hat{\mu}_D$, while P reveals information if and only if $\mu > \hat{\mu}_P$. Suppose these threshold strategies. Then, in case both D and P know μ, one of them reveals information. To see this, notice that if one of the agents reveals information, then $s = \mu$. If neither agent reveals information, then $s = E\left(\mu | l \leq \mu \leq \mu_x, m_D = m_P = \varnothing\right)$. If $\mu > E\left(\mu | l \leq \mu \leq \mu_x, m_D = m_P = \varnothing\right)$, P has an incentive to reveal information. If $\mu < E\left(\mu | l \leq \mu \leq \mu_x, m_D = m_P = \varnothing\right)$, D has an incentive to reveal information. The result that either D or P wants to reveal information implies that $\hat{\mu} = \hat{\mu}_P = \hat{\mu}_D$. Of course, my finding that either D or P wants to reveal information is a direct consequence of opposing preferences. Information that is good for i is bad for $-i$.

Let me now determine $\hat{\mu}$. At $\mu = \hat{\mu}$, each agent is indifferent between revealing μ and concealing μ. For D, this means

$$-\lambda_D - E\left(\mu | l \leq \mu \leq \mu_x, m_D = m_P = \varnothing\right) = -\lambda_D - \hat{\mu}$$

implying

$$\begin{aligned} \hat{\mu} &= \frac{\pi_P^\mu \left(1 - \pi_D^\mu\right) \frac{1}{2}\hat{\mu} + \pi_D^\mu \left(1 - \pi_P^\mu\right) \frac{1}{2}\left(\hat{\mu} + \mu_x\right) + \left(1 - \pi_D^\mu\right)\left(1 - \pi_P^\mu\right) \frac{1}{2}\mu_x}{1 - \pi_P^\mu \pi_D^\mu} \\ &= \frac{\left(1 - \pi_D^\mu\right) l + \left(1 - \pi_P^\mu\right) \mu_x}{2 - \pi_P^\mu - \pi_D^\mu} \qquad (1) \end{aligned}$$

Notice that if $\pi_P^\mu = \pi_D^\mu$, $\hat{\mu} = \frac{1}{2}\left(l + \mu_x\right)$. Then, the court imposes a "neutral" sentence in the absence of information. If $\pi_D^\mu > \pi_P^\mu$, in the absence of information, the court imposes a harsher sentence. If π_D^μ is high, the court tends to infer from $m_D = \varnothing$ that D wants to conceal information rather than that D does not possess information. The court will be more sceptical if $m_D = \varnothing$. As a result, D is more inclined to provide information. Likewise, if $\pi_P^\mu > \pi_D^\mu$, P has stronger incentives to reveal information than D. In the absence of information, the sentence will be mild.

I have now derived the threshold characterizing the agents' communication strategies for the case that $x_D = 1$. In a similar way, another threshold can be derived for the case that $x_D = 0$ and P has presented evidence that $x = 1$. In that case, the court knows that $\mu \in [\mu_x, h]$. Straightforward algebra shows that the equilibrium communication strategies for D and P are characterized by

$$\tilde{\mu} = \frac{\left(1 - \pi_D^\mu\right) \mu_x + \left(1 - \pi_P^\mu\right) h}{2 - \pi_P^\mu - \pi_D^\mu} \tag{2}$$

The thresholds defined by (1) and (2) characrterize the equilibrium strategies in the two communication games. Does any other equilibrium exists? The answer to this question is in the negative. Essential for the uniqueness of the equilibrium of each game is that there exists a positive probability that an agent is not able to present evidence to the court.[5] As a result, in any possible equilibrium we must have that $-h < E\left(\mu | m_P = \varnothing, m_D = \varnothing\right) < h$. From the agents' payoff functions it follows that both D and P follow a threshold strategy. Given that both D and P follow a threshold strategy, the thresholds derived above are unique.

Proposition 1 *Suppose an adversarial process in which the communication stage focuses on the sentence. Then, both D and P follow opposite threshold strategies characterized by the threshold* $\hat{\mu} = \frac{\left(1-\pi_D^\mu\right)l+\left(1-\pi_P^\mu\right)\mu_x}{2-\pi_P^\mu-\pi_D^\mu}$ *for* $x_D = 1$ *and by the threshold* $\tilde{\mu} = \frac{\left(1-\pi_D^\mu\right)\mu_x+\left(1-\pi_P^\mu\right)h}{2-\pi_P^\mu-\pi_D^\mu}$ *for* $x_D = 0$*: an informed D reveals information if* $\mu < \hat{\mu}$ *(or* $\mu < \tilde{\mu}$*), while an informed P reveals information if* $\mu > \hat{\mu}$ *(or* $\mu > \tilde{\mu}$*). If both D and P are informed, the court imposes the socially optimal sentence* $s = \mu$*. In case the court does not receive information about* μ*, the sentence is relatively mild if* $\pi_P^\mu > \pi_D^\mu$*, and relatively heavy if* $\pi_P^\mu > \pi_D^\mu$*. The equilibrium of each communication game is unique.*

The results presented in Proposition 1 are similar to the results derived by Shin (1994) who examines a persuasion game in which an arbitrator must make a decision on the basis of limited information provided by interested parties. He shows that when the parties have imperfect information, the arbitrator receives messages from more informed parties with greater scepticism.

Proposition 1 pertains to a situation in which the trial process does not focus on the question of being guilty or not guilty. This has already been established. This

[5] If $\pi_D^\mu = 1$ or $\pi_P^\mu = 1$, multiple equilibria do exist. In those cases, out of equilibrium beliefs are important. If the court is sceptical when receiving no information, then (1) and (2) still apply.

means that Proposition 1 is also relevant in civil trial processes that focus on the distribution of assets as in divorce cases. Then, s denotes what one party gets (and the other party does not get).

3.2 The Pleading Decision

Why would a defendant ever plead guilty? This section tries to give an answer to this question. As $x_D = 1$ leads to a certain conviction, it should at least increase the probability of a lower sentence. An implication is that D has no incentive to plead guilty if μ is high. This suggests that in equilibrium D follows a threshold strategy: plead guilty if $\mu < \mu_x$, and plead not guilty if $\mu \geq \mu_x$.

Let me now determine μ_x. Suppose that D anticipates equilibrium behavior in the communication stage as derived in the previous section. At $\mu = \mu_x$, D is indifferent between $x_D = 0$ and $x_D = 1$. Suppose $\mu = \mu_x$ and $x_D = 1$. Then, D's expected payoff equals

$$-\lambda_D - \pi_P^\mu \mu_x - (1 - \pi_P^\mu)\,\hat{\mu} \tag{3}$$

Equation (3) shows that pleading guilty leads to a certain conviction, but to a mild expected sentence. If $\mu = \mu_x$ and $x_D = 0$, then D's expected payoff equals

$$-\pi_P^x \lambda_D - \pi_P^x \pi_D^\mu \mu_x - \pi_P^x\,(1 - \pi_D^\mu)\,\tilde{\mu} \tag{4}$$

Pleading not guilty does not lead to a sure conviction but may result in a heavy sentence. Clearly, if pleading not guilty were to lead to a relatively mild expected sentence, then pleading guilty could not be part of an equilibrium. As a result, (4) implies that pleading guilty requires that the π_P^x is sufficiently high. Equating (3) and (4), using (2) and (1), yields

$$\mu_x = \frac{-\left(2 - \pi_P^\mu - \pi_D^\mu\right)\left(1 - \pi_P^x\right)\lambda_D - \left(1 - \pi_P^\mu\right)\left(1 - \pi_D^\mu\right) l + \pi_P^x \left(1 - \pi_D^\mu\right)\left(1 - \pi_P^\mu\right) h}{\left(1 - \pi_P^\mu \pi_D^\mu\right)\left(1 - \pi_P^x\right)} \tag{5}$$

D's pleading strategy is characterized by the threshold μ_x.

Proposition 2 *In an adversarial process, D pleads guilty if and only if $\mu < \mu_x$ with μ_x given by (5). An interior solution ($0 < \mu_x < h$) requires that π_P^x is sufficiently* *large, π_D^μ and π_P^μ are sufficiently small, and h is large relative to l and λ_D. The threshold μ_x is increasing in π_P^x and h, and decreasing in l and λ_D.*

Proposition 2 is the main result of this paper. In an adversarial system, the defendant may plead guilty. The analysis shows that defendants who would be given a mild sentence if all information would become available are most likely to plead guilty. Stubborn criminals, risking a severe sentence, do not plead guilty. As a result, pleading guilty leads to a mild sentence. The comparative static results presented in Proposition 2 are intuitive. As pleading guilty leads to a sure conviction, a higher cost of conviction (a higher value of λ_D) discourages pleading guilty. Likewise, if π_P^x is low, the probability of a conviction is small if the defendant pleads not guilty. As a result, the defendant is less likely to plead guilty. Finally, the higher is h, the more severe is the sentence when evidence is found on x, but not on μ. By pleading guilty, the defendant avoids such a severe sentence.

4 Welfare

In the previous section I have identified the conditions under which a guilty defendant pleads guilty. In this section, I examine the welfare consequences of the pleading decision. More specifically, I derive the optimal threshold, μ_x^s, from a social point of view, given D's and P's equilibrium strategies as derived in Section 3.1.

Let me first define a measure of social welfare. I assume that society wants to maximize the probability that a guilty defendant is convicted. Moreover, I assume that, if the court imposes a sentence, society attaches quadratic costs to deviations of s from μ. More specifically, my welfare criterion is

$$W^s(\mu_x) = \lambda_s \Pr(g = 1 | x = 1) - E\left[(s - \mu)^2 | x = 1\right] \tag{6}$$

where λ_s denotes the weight society attributes to the conviction objective relative to the sentence objective. I am aware that (6) is arbitrary and that numerous alternative specifications could be chosen. One advantage of the present specification is that it allows me to focus on the conviction decision and sentence decision separately. One reason why the welfare function can be kept simple is that in the present model innocent defendants are never convicted. Therefore, we do not have to specify the costs of convicting an innocent defendant.

Obviously, from a conviction point of view [the first term of (6)] the socially optimal value of $\mu_x = h$. Then, $\Pr(g = 1|x = 1) = 1$. From a sentence point of view, things are less clear. One can show, however, that $E\left[(s-\mu)^2|x=1\right]$ is maximized at $\mu_x = \frac{1}{2}(l+h)$.[6] This result clearly illustrates the signaling role of the pleading decision. In case the attorneys do not reveal μ, the court infers some information from the pleading decision. Clearly, if all guilty defendants were to plead guilty, the pleading decision would not contain any information about μ. All in all, I have found that the optimal threshold for the pleading decision is larger than the average sentence, $\mu_x^s > \frac{1}{2}(l+h)$. If λ_s is sufficiently small, $\mu_x^s < h$.

Proposition 3 *The socially optimal threshold for μ_x is higher than $\frac{1}{2}(l+h)$ and is smaller than h if λ_s is sufficiently small.*

Proof. One can write $E\left[(s-\mu)^2|x=1\right]$ as

$$\frac{\mu_x - l}{h-l}\left(\begin{array}{c} (1-d)\,p\int_l^{\left(\frac{(1-d)l+(1-p)\mu_x}{2-p-d}\right)} \frac{1}{\left(\frac{(1-d)l+(1-p)\mu_x}{2-p-d}\right)-l}\left(\left(\frac{(1-d)l+(1-p)\mu_x}{2-p-d}\right)-\mu\right)^2 d\mu+ \\ (1-p)\,d\int_{\left(\frac{(1-d)l+(1-p)\mu_x}{2-p-d}\right)}^{\mu_x} \frac{1}{\mu_x-\left(\frac{(1-d)l+(1-p)\mu_x}{2-p-d}\right)}\left(\left(\frac{(1-d)l+(1-p)\mu_x}{2-p-d}\right)-\mu\right)^2 d\mu+ \\ (1-p)(1-d)\int_l^{\mu_x}\frac{1}{\mu_x-l}\left(\left(\frac{(1-d)l+(1-p)\mu_x}{2-p-d}\right)-\mu\right)^2 d\mu \end{array}\right)+$$
$$\frac{h-\mu_x}{h-l}\left(\begin{array}{c} (1-d)\,p\int_{\mu_x}^{\left(\frac{(1-d)\mu_x+(1-p)h}{2-p-d}\right)} \frac{1}{\left(\frac{(1-d)\mu_x+(1-p)h}{2-p-d}\right)-\mu_x}\left(\left(\frac{(1-d)\mu_x+(1-p)h}{2-p-d}\right)-\mu\right)^2 d\mu+ \\ (1-p)\,d\int_{\left(\frac{(1-d)\mu_x+(1-p)h}{2-p-d}\right)}^{h} \frac{1}{h-\left(\frac{(1-d)\mu_x+(1-p)h}{2-p-d}\right)}\left(\left(\frac{(1-d)\mu_x+(1-p)h}{2-p-d}\right)-\mu\right)^2 d\mu+ \\ (1-p)(1-d)\int_{\mu_x}^{h}\frac{1}{h-\mu_x}\left(\left(\frac{(1-d)\mu_x+(1-p)h}{2-p-d}\right)-\mu\right)^2 d\mu \end{array}\right)$$

Differentiating with respect to μ_x yields the first-order condition

$$\left(1-\pi_P^\mu\pi_D^\mu\right)\left(1-\pi_D^\mu\right)\left(1-\pi_P^\mu\right)\frac{h+l-2\mu_x}{\left(2-\pi_D^\mu-\pi_P^\mu\right)^2}=0$$

implying $\mu_x^s = \frac{1}{2}(l+h)$. ■

[6] My assumption that the distribution of μ is symmetric is responsible for the result that exactly half of the guilty defendants should plead guilty. Important is that from a sentence point of view, it is optimal that only part of the guilty defendants plead guilty.

In practice, guilty defendants may have too weak or too strong incentives to plead guilty, $\mu_x^s < \mu_x^*$. Lower (higher) sentences may encourage (discourage) defendants to plead guilty. For example, the court may impose, say, a third reduction of the sentence if the defendant pleads guilty.

5 Two extensions

So far, in my model innocent defendants could not be convicted because of the assumption that P never finds evidence against an innocent defendant. In reality, innocent people are sometimes convicted. It is easy to adapt my model to allow for the possibility that also innocent people may be convicted. In the model of Section 2, the prosecutor could only find evidence that the defendant committed the crime if the defendant is guilty. To allow for the possibility that innocent defendants can be convicted one may assume that with probability $\pi_P^{x,i}$, P finds hard evidence that an innocent suspect is guilty. As to exogenous circumstances, μ, we may assume that with probability $\pi_P^{\mu,i}$, P finds hard evidence on μ in case of an innocent suspect, and with probability $\pi_D^{\mu,i}$, D finds hard evidence on μ in case of an innocent suspect. Clearly, in a model like this an innocent defendant may also plead guilty. In fact, an innocent suspect faces the same kind of trade-off as a guilty suspect: Pleading guilty leads to a certain conviction, but to a lower expected sentence. Again pleading guility requires that $\pi_P^{x,i}$ is sufficiently large. The idea that society attaches high costs to innocent people being convicted is widespread. From this point of view, it is socially undesirable that innocent people plead guilty and are convicted. However, as long as the defendant's choice to plead guilty is voluntary, pleading guilty does not conflict with the defendant's interest. Another implication of allowing for the possibility that P finds hard evidence on an innocent suspect is that the court may infer information about x from the pleading decision. In case innocent suspects are more likely to plead guilty than guilty suspects (e.g. because mitigating circumstances especially apply to innocent suspects), the court may get reluctant to convict suspects solely on the basis of guilty pleas.

A hidden cost of the possibility to plead guilty is that it may reduce deterrence. It is possible to add a stage to the basic model in which people decide on whether or not to commit an illegal act. People vary in their inclination to commit a crime.

In line with the spirit of my model one may assume that persons for whom s is low are the least prone to commit an illegal act. Then, modelling deterrence amounts to explaining l. One of my main results is that for a defendant to whom mitigating circumstances apply, the benefit of a lower sentence may compensate for the cost of a certain conviction. This also means that if guilty pleas occur in equilibrium (μ_x has an interior solution), the possibility to plead guilty may encourage people whose s is low to commit crimes. Hence, a possible cost of guilty pleas is less deterrence.

6 Conclusion

In this paper I have examined the role of the pleading decision in criminal trials. I have shown that the pleading decision may serve as a signaling device differentiating defendants who should be punished harsly from defendants who should be punished mildly. Defendants pleading guilty increase conviction rates. If only guilty defendants plead guilty this is good from a social perspective. Another function of a judicial system is imposing proper sentences. Plea decisions contribute to this function optimally if half of the guilty defendants plead guilty.

7 References

Austen Smith, David, 1994, Strategic Transmission of Costly Information, *Econometrica*, 62, 955-963.

Baker, Scott and Claudio Mezzetti, 2001, Prosecutorial Resources, Plea Bargaining, and the Decision to Go to Trial, *Journal of Law, Economics and Organization*, 17, 149-167.

Dewatripont, Mathias, and Jean Tirole, 1999, Advocates,*Journal of Political Economy*,107, 1-39.

Dur, Robert A.J. and Otto H. Swank, 2004, Producing and Manipulating Information, *Economic Journal*, 115, 185-199.

Gordon, Sanford C., and Gregory A. Huber, 2009, The Political Economy of Prosecution, *Annual Review of Law and Social Science*, 5, 135-156.

Grossman, Gene M., and Micheal L. Katz, 1983, Plea Bargaining and Social Welfare, American Economic Review, 73, 749-757.

Kim, Chulyoung, 2010a, The Value of Information in Legal Systems, Discussion paper, University of California, San Diego.

Kim, Chulyoung, 2010b, Partisan Advocates, Discussion paper, University of California, San Diego.

Milgrom, Paul, and John Roberts, 1986, Relying on Information of Interested Parties, *RAND Journal of Economics*,17, 18-32.

Posner, Richard A., 1999, An Economic Approach to the Law of Evidence, Stanford Law Review, 51, 1477-1546.

Reinganum, Jennifer F., 1988, Plea Bargaining and Prosecutorial Discretion, American Economic Review, 78, 713-728.

Shin, Hyon Song, 1994, The Burden of Proof in an Game of Persuasion, *Journal of Economic Theory*, 64, 253-264.

Sobel, Joel, 1985, Disclosure of evidence and resolution of disputes: Who should bear the burden of proof? In Alvin E. Roth, ed., *Game Theoretic Models of Bargaining* Cambridge, chapter 16, Cambridge University Press.